Ivan John Naparota, March 24, 2018

II. Model 1: **Poisson Distribution**

Formula: $\boxed{P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \text{for } k = 0, 1, \ldots}$

When is the Poisson distribution an appropriate model? [3]

- The set of raw data are a set of **discrete random variables**
- **k is the number of times an event occurs in an interval**, in the gathered data, we have the arrival of students in a particular class with values 39, 36, 34, and so on.
- **The occurrence of one event does not affect the probability that a second event will occur**, the attendance today will not influence tomorrow's attendance.
- **Two events cannot occur at exactly the same instant**

Parameters for poisson distribution,

Mean $\lambda = \frac{Summation\ of\ K's}{sample\ size}$

$= \frac{39 + 36 + 34 ..}{54}$

= 36.7593

e = 2.718281828 (euler's number)

k = can take any values from 0,1,2,3 . . . but in the gathered data, it is very particular that the values are 39, 36, 34, and so on.

Model 2: **Binomial Distribution**

Formula: $$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ for } k = 0, 1, \ldots, n.$$

Use of the binomial distribution requires three assumptions: [2]

- Each replication of the process results in one of two possible outcomes (success or failure), **in this case, we can consider the two possibilities to be to attend the class or to absent.**
- The probability of success is the same for each replication, (solution for probability presented below) and
- The replications are independent, meaning here that a success in one patient does not influence the probability of success in another, **for our case, let us consider that the attendance of a single student does not depend on his classmate's attendance.**

Parameters for Binomial Distribution,

N = number of students in a class = 40

P = $\frac{Summation\ of\ all\ Attendance\ Samples}{54 trials (40 students)}$

P = $\frac{39+36+34+\ldots}{54 trials (40 students)}$

P = 0.9189814815

Results and Discussions:

Summary:

	POISSON		BINOMIAL	
X	Expected	Observed	Expected	Observed
33	3.065248	2	1.419501	2
34	3.314011	4	3.314953	4
35	3.480592	5	6.445892	5
36	3.554003	13	10.15484	13
37	3.530883	12	12.45242	12
38	3.4156	8	11.151	8
39	3.21936	8	6.48637	8
40	2.958536	2	1.839349	2

POISSON DISTRIBUTION

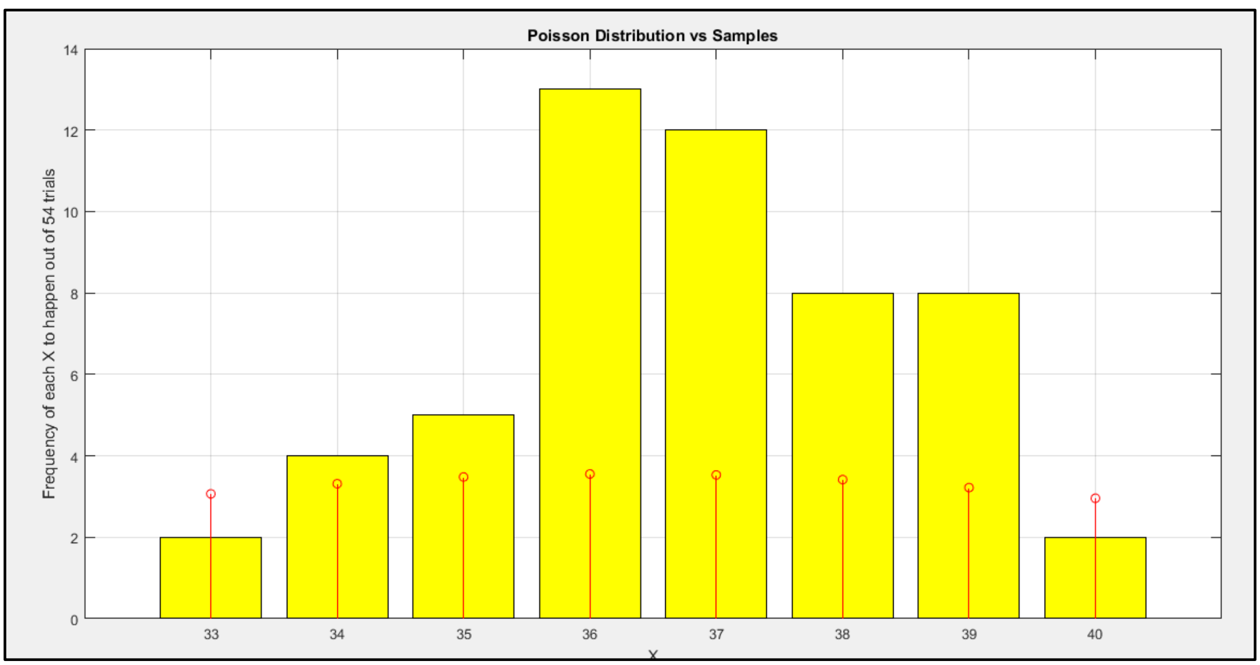


Figure 1 Poisson Distribution vs Samples

Red = Poisson Distribution

Yellow = Samples

This is the graph of the “Poisson Distribution vs The Samples”, in the y-axis are the occurrence of the different attendances out of 54 trials and the x – axis is the attendance of the students, intuitively, just by looking at the graph we can already say that this model will give us significant difference in its output (the occurrence of different attendances out of 54 trials) with respect to our sample variables, thus making this model a bad fit for our samples, nevertheless, let us test its goodness of fit using the Pearson Chi – Square goodness of fit test.

BINOMIAL DISTRIBUTION

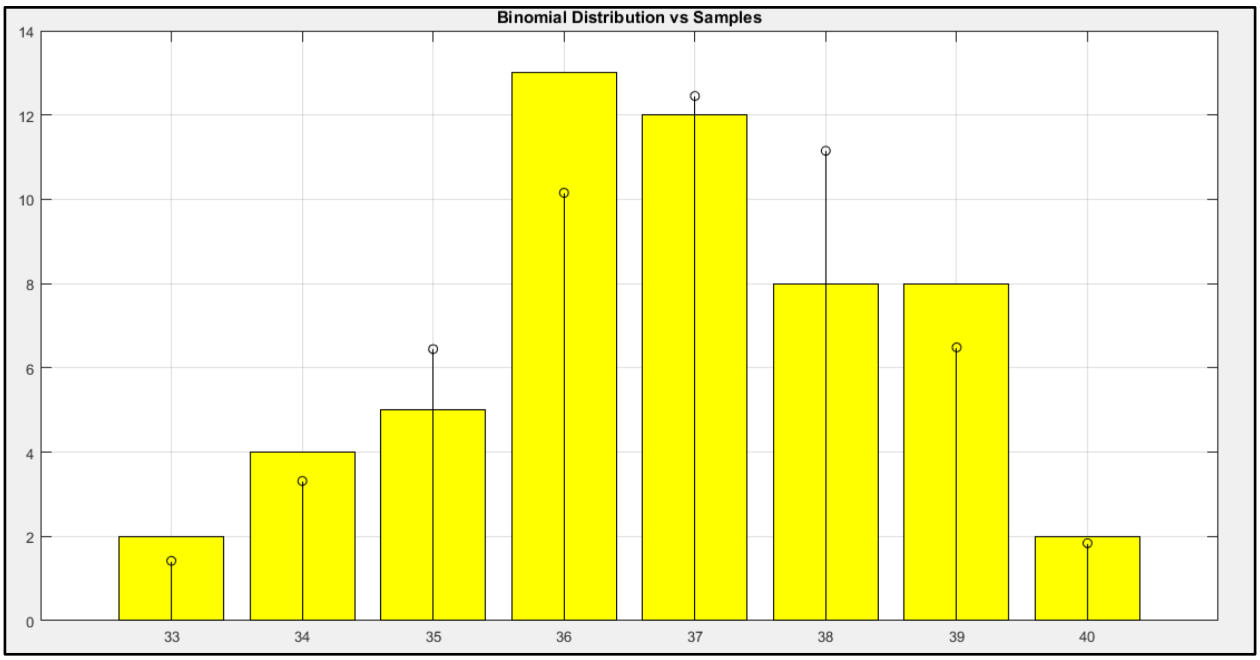


Figure 2 Binomial Distribution vs Samples

Yellow = samples

Black = Binomial Distribution

The figure above is the frequencies of having the different values of attendances (33,34,35…40) out of 54 trials of the Binomial Distribution model and the Samples. It is observable that this model gives a closer output than the previous model, but still, let us test the model’s goodness of fit with Chi Square.

Testing the models

Chi – Square Formula:

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where:

O = is the observed data (Samples)

E = expected data (Poisson)

Chi – square test considers these hypotheses:

Ho = there is no significant difference in the expected and observed data

Ha = there is significant difference in the expected and observed data

Parameters:

Degrees of freedom = number of categories – 1

Degrees of freedom = 8 – 1 = **7**

Alpha = **0.05,** "The minimum probability for rejecting a null hypothesis in the sciences is generally 0.05" [1]

Solving for Chi – Square,

	POISSON		BINOMIAL	
X	Expected	Observed	Expected	Observed
33	3.065248	2	1.419501	2
34	3.314011	4	3.314953	4
35	3.480592	5	6.445892	5
36	3.554003	13	10.15484	13
37	3.530883	12	12.45242	12
38	3.4156	8	11.151	8
39	3.21936	8	6.48637	8
40	2.958536	2	1.839349	2

For poisson:

$$X^2 = \frac{(2-3.065248)^\wedge 2}{3.065248} + \frac{(4-3.314011)^\wedge 2}{3.314011} + \ldots$$

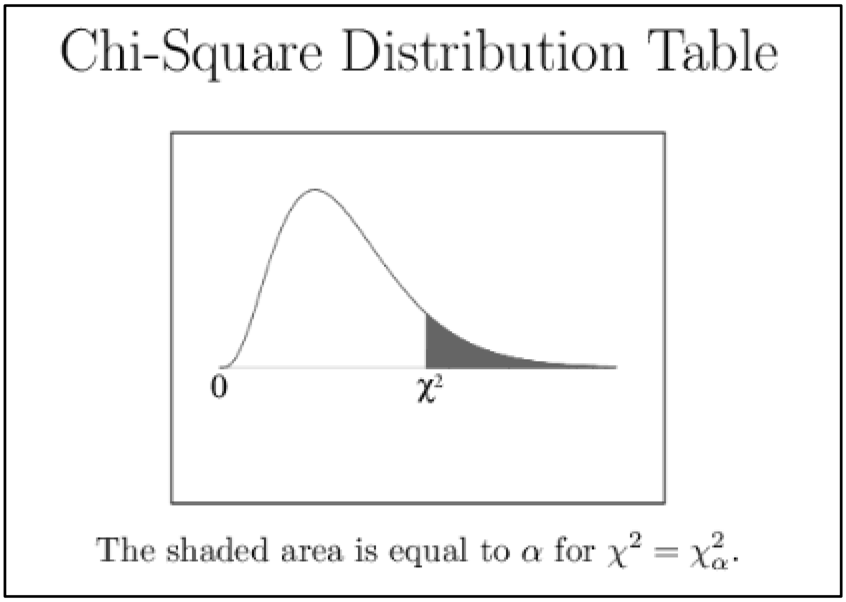
$\underline{X^2 = 60.1582}$

For Binomial Distribution:

$$X^2 = \frac{(2-1.419501)^\wedge 2}{1.419501} + \frac{(4-3.314953)^\wedge 2}{3.314953} + \ldots$$

$\underline{X^2 = 2.7745}$

Analysing the Results:



Critical Values of Chi Square

df	0.05	0.01	df	0.05	0.01
1	3.84	6.64	16	26.30	32.00
2	5.99	9.21	17	27.59	33.41
3	7.82	11.34	18	28.87	34.80
4	9.49	13.28	19	30.14	36.19
5	11.07	15.09	20	31.41	37.57
6	12.59	16.81	21	32.67	38.93
7	14.07	18.48	22	33.92	40.29
8	15.51	20.09	23	35.17	41.64
9	16.92	21.67	24	36.42	42.98
10	18.31	23.21	25	37.65	44.31
11	19.68	24.72	26	38.88	45.64
12	21.03	26.22	27	40.11	46.96
13	22.36	27.69	28	41.34	48.28
14	23.68	29.14	29	42.56	49.59
15	25.00	30.58	30	43.77	50.89

For Poisson Distribution:

Referring to the Chi Square table with alpha = 0.05 and df of 7, we can get a Chi Square value of 14.07 but our computation of Chi Square gives us a result way beyond that which is 60.1582, let us take a look at the Chi Square curve, the shaded part are the Chi Square values that dictate the rejection of null hypothesis (Ho), it is pretty obvious that our computed

Chi Square which is approximately 60 is far beyond that 14.07 mark, this tells us that we should **reject our null hypothesis (Ho). This means that there is a significant difference in our expected and observed data.**

For Binomial Distribution:

The computed Chi square for the Binomial Distribution model provides us an answer of 2.7745 which is located at the left side of the shaded area, this calls for the **acceptance of the null hypothesis. This means that there is no significant difference in our expected and observed data.**

Conclusion:

Based from various computations and results, out of the two models made, **the Binomial Distribution is the best model to describe the behaviour of the students' attendance in a particular class.**

References:

1. https://sites.google.com/a/cjeagles.org/mrs-ooten-s-courses/home/ap-biology/genetics/chi-square-analysis

2. http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Probability/BS704_Probability7.html

3.https://en.wikipedia.org/wiki/Poisson_distribution#Assumptions:_When_is_the_Poisson_distribution_an_appropriate_model?